UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

DEPARTAMENTO DE ESTADÍSTICA E INFORMÁTICA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

TRABAJO INTEGRADOR

Inferencia Estadística 2025-II

Integrante	Código
Montúfar Paiva Yeraldi Mercedes	20230400
Kay Daniela L. Zavala Malpartida	20230420
Rojas Taco Fabiana	2023
Castillo Ruiz Mauricio Gabriel	20230384
Gómez Vigo Héctor Estefano	20230397
Villanueva Huamani Alexander Ruben	20230419

Docente: Fernando Miranda Villagómez

LA MOLINA - LIMA - PERÚ 2025

Distribución Pareto:

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}} I_{(\alpha,\infty)}(x)$$
$$E(X) = \frac{\alpha\beta}{\beta-1} \quad ; \quad \text{VAR}(X) = \left(\frac{\alpha}{\beta-1}\right)^2 \frac{\beta}{\beta-2}$$

Pregunta 1

1.1 EMV para parametro alfa:

Usaremos el Método de máxima verosimilitud para poder hallar el primer estimador para alfa:

$$L(\alpha,\beta) = \prod_{i=1}^{n} f(x_i;\alpha,\beta) = \prod_{i=1}^{n} \frac{\beta\alpha^\beta}{x_i^{\beta+1}} I_{[\alpha,\infty)}(x_i) \quad = \beta^n\alpha^{n\beta}\left(\prod_{i=1}^{n} x_i^{-(\beta+1)}\right) \times \prod_{i=1}^{n} I_{[\alpha,\infty)}(x_i)$$

Analizando la indicadora:

$$\begin{aligned}\prod_{i=1}^{n} I_{[\alpha,\infty]}(x_i) &= I\left(\bigcap_{i=i}^{n}\{x_i \geq \alpha\}\right)\\ &= I(\alpha \leq x_1, \alpha \leq x_2, \ldots \alpha \leq x_n)\\ &= I(\alpha \leq mi(x_1,\ldots,x_n\}) = I(\alpha \leq y_1]\end{aligned}$$

donde y1 = X(1) es el mínimo de la muestra:

$$El \text{ EMV de } \theta = \alpha \text{ es } \; T = Y_1.$$

1.1.2 Propiedades:

a) Insesgamiento: Distribución del mínimo:

$$X_1 \ldots X_n \sim \text{ Pareto } (\alpha,\beta) \text{ i i d}$$
$$F(x) = 1 - \left(\frac{\alpha}{x}\right)^\beta, x \geqslant \alpha$$
$$y = x_i$$
$$Fx_i(y) = n[1-F(y)]^{n-1}\mathcal{F}(y) \to n\left[\left(\frac{\alpha}{\beta}\right)^\beta\right]^{\beta-1} \cdot \frac{\beta\alpha^\beta}{y^{\beta+1}}$$
$$= n\left[\frac{\alpha^{\beta(n-1)}}{\alpha^{\beta(n-1)}}\right] \cdot \frac{\beta\alpha^\beta}{y^{\beta+1}} = \frac{(n\beta)\alpha^{n\beta}}{y^{n\beta+1}}, y \geqslant \alpha$$

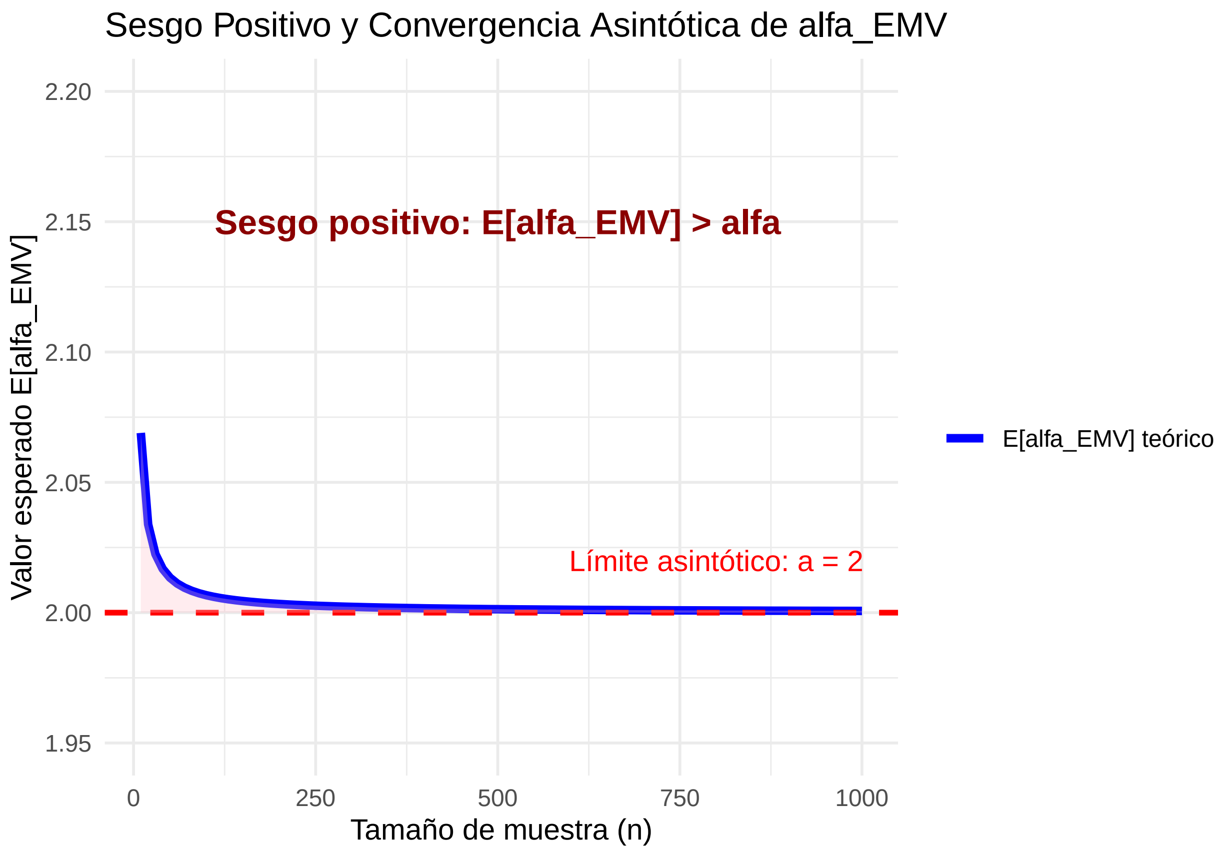
$$\to x_1 \sim \text{Pareto}(\alpha, n\beta)$$

Hay sesgo positivo, es decir que sobrestima al parametro, no cumple propiedades de Insesgamiento. Ahora lo veremos mejor con los valores para nuestra data:

```
# 1. Valor observado del mínimo
alpha_EMV_observado <- min(reclamos)
# 2. Valor esperado teórico del mínimo
E_alpha_EMV_teorico <- (n * beta_true * alpha_true) / (n * beta_true - 1)
```

```
## [1] "El valor observado es: 2.00031032898679 y el valor esperado es: 2.00066688896299"
```

El valor observado en los datos fue de 2.00031, mientras que el valor esperado teórico es 2.00067, ambos ligeramente por encima del valor real del parámetro (alfa = 2.00000). Esto confirma que, para muestras finitas, el estimador tiende a sobrestimar el parámetro verdadero, aunque en este caso la magnitud del sesgo es muy pequeña (del orden de 0.0003 a 0.0007). La cercanía entre el valor observado y el teórico valida la expresión matemática del sesgo y respalda la propiedad de insesgamiento asintótico, ya que a medida que el tamaño de muestra aumenta, el sesgo tiende a cero.



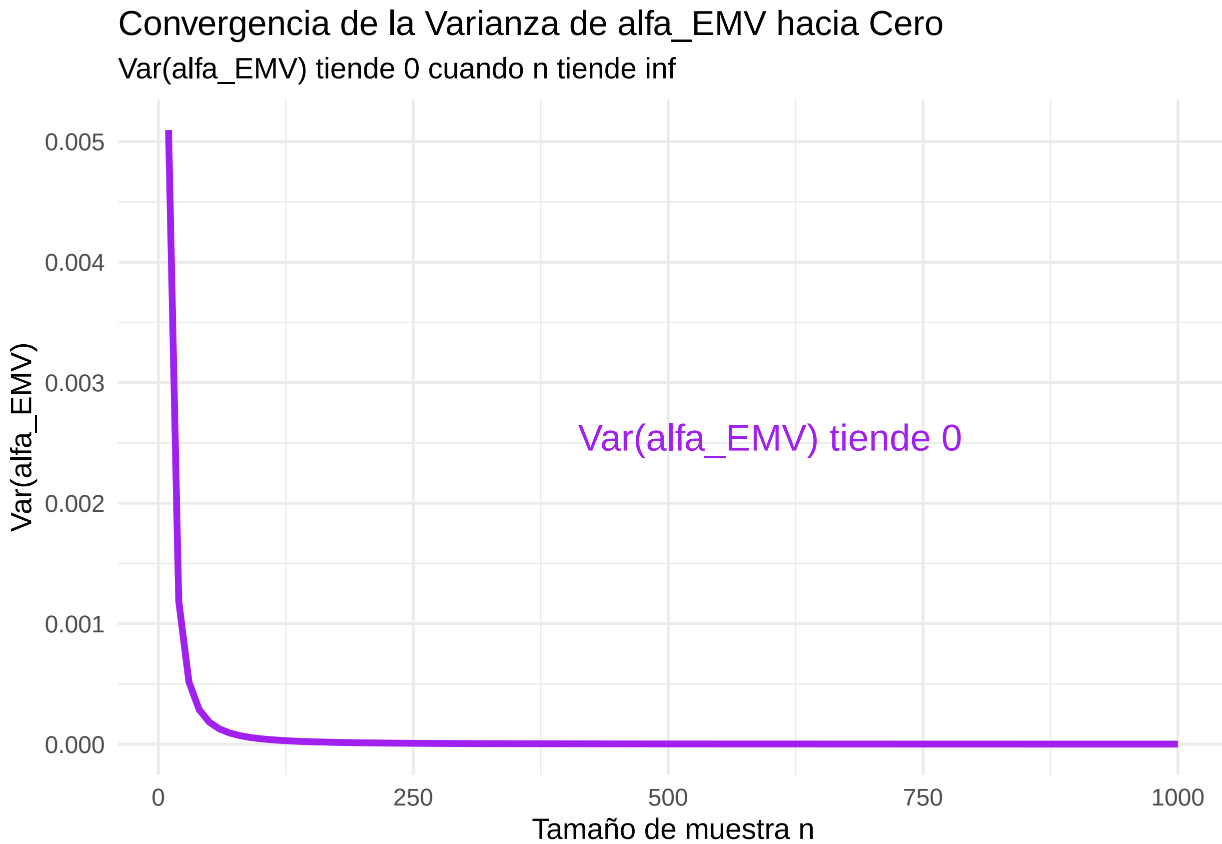
de nuestra base de datos, tomamos submuestras que van incrementando su valor para verificar que tendiendo al infinito (tamaños suficientemente grande) podemos ver como el valor tiende al parámetro de alfa = 2.

b) Consistencia: Podemos hacer uso del Teorema 2, por la propiedad anterior del insesgamiento asintotico

- $\lim_{n\to\infty} E(y_1) = \alpha$ ya que $\lim_{n\to\infty} \dfrac{n\beta \cdot \alpha}{n\beta - 1} = \lim_{n\to\infty} \dfrac{\beta\alpha}{\beta - \frac{1}{n}} = \dfrac{\beta\alpha}{\beta} = \alpha$
- $\lim_{n\to\infty} \text{Var}(y_1) = \lim_{n\to\infty} \dfrac{n\beta\alpha^2}{(n\beta-1)^2(n\beta-2)} = \lim_{n\to\infty} \dfrac{n\beta\alpha^2}{n^3\beta^3} = \lim_{n\to\infty} \dfrac{\alpha^2}{n^2\beta^2} = 0$

Como podemos obervar cumple ambas condiciones, por ende decimos que nuestro estimador es consitente n otras palabras el estimador converge en probabilidades al parametro.

** Gráfico:** Hemos visto ya un gráfico para solidar la primera demostración, ahora veamos uno que consolide la segunda condición:



c) Suficiencia:

$$P(x = x/t = T) = \frac{P\left(x_1 = x_1 \ldots x_n = x_n, T = y_1\right)}{P(T = y_1)}$$

$$\frac{\left(\frac{\beta\alpha^\beta}{x_1^{\beta+1}}\right)\left(\frac{\beta\alpha^\beta}{x_2^{\beta+1}}\right)\cdots\left(\frac{\beta\alpha^\beta}{x_n^{\beta+1}}\right)}{\frac{n\beta\alpha^{n\beta}}{y^{n\beta+1}}}, \quad x_i \geqslant \alpha,\ i = 1, 2, \ldots, n$$

$$= \frac{\beta^n \cdot \alpha^{n\beta} \cdot \prod_{i=1}^n x_i^{-(\beta+1)}}{\frac{n\beta\alpha^{n\beta}}{y^{n\beta+1}}} = \frac{\beta^{n-1} \cdot y^{n\beta+1} \cdot \prod_{i=1}^n x_i^{-(\beta+1)}}{n}$$

Dada la dsitribución condicional dado Y = X1, no depende de alfa, entonces podemos decir que es una estadística suficiente para alpha, cumple la propiedad.

d) Ancilaridad: La distribución del mínimo $x_{(1)}$ es:

$$x_{(1)} \sim \text{Pareto}(\alpha, n\beta)$$

$$f_{x_{(1)}}(t) = \frac{n\beta\alpha^{n\beta}}{t^{n\beta+1}}, \quad t \geqslant \alpha$$

Como vemos su distribución depende de alfa, no cumple con la propiedad de ancilaridad.

e) Completitud: Calculamos la esperanza de $g(T)$:

$$E[g(T)] = \int_{\alpha}^{\infty} g(t) \cdot \frac{n\beta\alpha^{n\beta}}{t^{n\beta+1}}\, dt.$$

Exigimos que $E[g(T)] = 0$ para todo $\alpha > 0$:

$$\int_{\alpha}^{\infty} \frac{g(t)}{t^{n\beta+1}}\, dt = 0 \quad \forall\alpha > 0.$$

Derivamos ambos lados respecto a α:

$$\frac{d}{d\alpha}\left[\int_{\alpha}^{\infty} \frac{g(t)}{t^{n\beta+1}}\, dt\right] = -\frac{g(\alpha)}{\alpha^{n\beta+1}} = 0.$$

Esto implica que:

$$g(\alpha) = 0 \quad \forall\alpha > 0.$$

Como α es cualquier valor positivo, concluimos que $g(t) = 0$ para todo t en el soporte de T. Por lo tanto, $T = X_{(1)}$ es una **estadística completa** para α cuando β es conocido.

Por el Teorema de Bahadur, como es suficiente y completo, podemos afirmar que nuestro estimador es minimal suficiente.

f) Optimalidad: En nuestro caso alfa_EMV no es óptimo porque es sesgado. Pero si existe un estimador óptimo basado en la estadística suficiente que se podría obtener mediante corrección del sesgo. (Teorema de Lehman-Scheffé)

1.2 EMV para parametro Beta:

Para hallar el estimador debemos remplazar nuestro primer estimador para el parametro alfa:

$$\hat{\alpha}_{MLE} = \min\{X_1, X_2, \ldots, X_n\}$$

Como sabemos el máximo en L(alpha,beta) es igual al máximo de Ln(Lalpha,beta)) entonces haremos uso del ln:

$$L(\alpha, \beta) = \beta^n \cdot \alpha^{n\beta} \cdot \textstyle\prod_{i=1}^{n} X_i^{-(\beta+1)}$$
$$\ln(L(\alpha, \beta)) = n \cdot \ln(\beta) + n\beta\ln(\alpha) - (\beta + 1)\textstyle\sum_{i=1}^{n} \ln(X_i)$$

$$
\begin{aligned}
\frac{d\ln(L(\alpha,\beta))}{d\beta} &= \frac{n}{\beta} + n\ln\alpha - \sum_{i=1}^{n}\ln X_i \\
&= \frac{n}{\beta} + n\ln\alpha - \sum_{i=1}^{n}\ln X_i = 0 \\
&\Rightarrow \frac{n}{\beta} = \sum_{i=1}^{n}\ln X_i - n\ln\alpha \\
&\Rightarrow \frac{n}{\beta} = \sum_{i=1}^{n}\ln\left(\frac{X_i}{\alpha}\right) \\
&\Rightarrow \hat{\beta} = \frac{n}{\sum_{i=1}^{n}\ln\left(\frac{X_i}{x_1}\right)}
\end{aligned}
$$

1.2.1 Propiedades:

a) Insesgamiento Primero analizaremos una parte del denominador para determinar su distribución y se facilite el procedimiento.

$$
\begin{aligned}
F_Y(y) = P(Y_i \leq y) &= P\left(\ln\left(\frac{X_i}{\alpha}\right) \leq y\right) = P(X_i \leq \alpha e^y) \\
&= 1 - \left(\frac{\alpha}{\alpha e^y}\right)^{\beta} = 1 - e^{-\beta y}, \quad y \geq 0 \\
f_Y(y) &= \frac{d}{dy}F_Y(y) = \beta e^{-\beta y}, \quad y \geq 0
\end{aligned}
$$

Analizando el resultado de la distribución, notamos que tiene la forma de la exponencial. Entonces el mínimo igual será exponencial con parametro n por beta. Ademas tomamos a T como la sumatoria, donde por propiedad de ln de una división podemos desplegarlo.

$$
\begin{aligned}
Y_i &\sim \text{Exponencial}(\beta) \\
W = Y_{(1)} &= \min(Y_1, \ldots, Y_n) \sim \text{Exp}(n\beta) \\
T &= \sum_{i=1}^{n}(Y_i - Y_{(1)}) \sim \text{Gamma}(n-1, \beta) \\
\hat{\beta} &= \frac{n}{T}, \quad T \sim \text{Gamma}(n-1, \beta) \\
E\left[\frac{1}{T}\right] &= \frac{\beta}{n-2}, \quad n > 2 \\
E[\hat{\beta}] &= n \cdot E\left[\frac{1}{T}\right] = \frac{n\beta}{n-2}
\end{aligned}
$$

Acá vemos que hay un sesgo presente

$$
\text{Sesgo} = E[\hat{\beta}] - \beta = \frac{n\beta}{n-2} - \beta = \frac{2\beta}{n-2}
$$

Por ende, no es un estimador insesgado. Ademas al ser positivo podemos decir que sobrestima al parametro.

1.3 Estimador para la Media Poblacional (μ)

Consideramos la media muestral como estimador natural del valor esperado de la distribución:

$$\bar{X} = T(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Para que los momentos existan en una distribución Pareto, debemos asumir que $\beta > 1$ (para la esperanza) y $\beta > 2$ (para la varianza). El parámetro a estimar es:

$$\mu = E(X) = \frac{\beta\alpha}{\beta - 1}$$

1.3.1 Propiedades:

a) Insesgamiento: Calculamos el valor esperado del estimador:

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\ &= \frac{1}{n} \cdot n \cdot \left(\frac{\beta\alpha}{\beta-1}\right) = \frac{\beta\alpha}{\beta-1} = \mu \end{aligned}$$

El estimador es **estrictamente insesgado** para μ. Verificamos esto con la base de datos:

```
# 1. Valor observado de la media muestral
mu_muestral_obs <- mean(reclamos)

# 2. Valor esperado teórico (mu)
mu_teorico <- (beta_true * alpha_true) / (beta_true - 1)
```

```
## [1] "La media muestral observada es: 2.96908 y el valor esperado teórico es: 3"
```

b) Consistencia: Para demostrar la consistencia, verificamos las condiciones del Teorema de Chebyshev (asumiendo $\beta > 2$):

1. **Insesgadez:** $\lim_{n\to\infty} E(\bar{X}) = \mu$ (ya demostrado).
2. **Varianza tiende a cero:**

$$\begin{aligned} \mathrm{Var}(\bar{X}) &= \frac{\mathrm{Var}(X)}{n} \\ &= \frac{1}{n}\left[\left(\frac{\alpha}{\beta-1}\right)^2 \frac{\beta}{\beta-2}\right] \end{aligned}$$

$$\lim_{n\to\infty} \mathrm{Var}(\bar{X}) = \lim_{n\to\infty} \frac{C}{n} = 0$$

Al cumplirse ambas condiciones, $\bar{X}$ es un estimador **consistente** para μ.

c) Suficiencia: Aplicamos el Teorema de Factorización de Fisher-Neyman a la función de verosimilitud:

$$L(\alpha, \beta) = \underbrace{\beta^n \alpha^{n\beta} \cdot I_{(\alpha,\infty)}(x_{(1)})}_{g(x_{(1)},\alpha,\beta)} \cdot \underbrace{\left(\prod_{i=1}^{n} x_i\right)^{-(\beta+1)}}_{h(x_1,\ldots,x_n)}$$

Como se observa, los estadísticos suficientes conjuntos para (α, β) son el mínimo $X_{(1)}$ y el producto $\prod X_i$ (o equivalentemente $\sum \ln X_i$). La media muestral $\bar{X} = \frac{1}{n} \sum X_i$ **no puede factorizarse** de manera que contenga toda la información de los parámetros.

Por lo tanto, $\bar{X}$ **no es un estadístico suficiente** para los parámetros de la distribución Pareto.

d) Ancilaridad: Un estadístico es ancilar si su distribución no depende de los parámetros. La distribución de $\bar{X}$ para una Pareto no tiene una forma cerrada sencilla (es una suma de variables Pareto), pero su esperanza $E(\bar{X}) = \frac{\alpha\beta}{\beta-1}$ y su varianza dependen directamente de α y β.

Al depender sus momentos (y por ende su distribución) de los parámetros, el estimador **no es ancilar**.

e) Completitud: Dado que $\bar{X}$ no es un estadístico suficiente para la familia Pareto, no se suele analizar su completitud como estimador. Sin embargo, sabemos que el estadístico suficiente conjunto $(X_{(1)}, \sum \ln X_i)$ es completo, pero la suma aritmética $\sum X_i$ no lo es bajo esta estructura de familia no exponencial (en el sentido de los parámetros naturales de la Pareto).

f) Optimalidad: Un estimador es óptimo (UMVUE) si es insesgado y su varianza alcanza la Cota Inferior de Cramér-Rao (CICR) o si es función de un estadístico suficiente y completo.

1. **Eficiencia:** Al no ser función del estadístico suficiente $(X_{(1)}, \sum \ln X_i)$, la media muestral pierde información.
2. **Comparación:** Existe otro estimador para μ, basado en los EMV de α y β ($\hat{\mu} = \frac{\hat{\beta}\hat{\alpha}}{\hat{\beta}-1}$), que asintóticamente tiene menor varianza que $\bar{X}$.

Conclusión: $\bar{X}$ **no es un estimador óptimo** para la media de una población Pareto, aunque sea fácil de calcular e insesgado.

1.5 Estimador : Estimador Insesgado de Varianza Uniformemente Mínima (EIVUM) para el parámetro β

Dado que en la sección anterior se determinó que el Estimador de Máxima Verosimilitud (EMV) para β es sesgado, procederemos a obtener el estimador óptimo aplicando el **Teorema de Lehmann-Scheffé**.

Según la teoría de la optimalidad, si logramos construir un estimador insesgado que sea función de una estadística suficiente y completa, dicho estimador será el **Estimador Insesgado de Varianza Uniformemente Mínima (EIVUM)**.

El procedimiento se detalla en los siguientes cuatro pasos.

Paso 1: Identificación de la Estadística Suficiente y Completa

Analizando la función de densidad de la distribución Pareto, observamos que pertenece a la familia exponencial k-paramétrica respecto al parámetro de forma β.

Bajo la condición de que el parámetro de escala α es desconocido y estimado mediante el mínimo muestral $X_{(1)}$, la teoría de suficiencia indica que la estadística T que captura toda la información sobre β es la suma de los logaritmos de las razones muestrales:

$$T = \sum_{i=1}^{n} \ln\left(\frac{X_i}{X_{(1)}}\right)$$

Esta estadística T es **suficiente y completa** para el parámetro β, condición necesaria para aplicar el teorema de Lehmann-Scheffé.

Paso 2: Distribución Muestral de la Estadística

A partir de las propiedades de la transformación de variables aleatorias Pareto, se deduce que la estadística T sigue una distribución Gamma.

Dado que se ha estimado un parámetro adicional (α), los grados de libertad se ajustan a $(n-1)$. Por tanto:

$$T \sim \text{Gamma}(n-1, \beta)$$

Paso 3: Verificación del Sesgo y Corrección (Método de la Esperanza)

Partimos del Estimador de Máxima Verosimilitud hallado previamente:

$$\hat{\beta}_{EMV} = \frac{n}{T}$$

Para verificar si es insesgado, calculamos su valor esperado $E[\hat{\beta}_{EMV}]$. Utilizando la propiedad de la esperanza inversa para una variable con distribución Gamma(k, λ), donde:

$$E\left[\frac{1}{T}\right] = \frac{\lambda}{k-1},$$

se obtiene:

$$E\left[\frac{1}{T}\right] = \frac{\beta}{(n-1)-1} = \frac{\beta}{n-2}$$

Sustituyendo en la esperanza del EMV:

$$E[\hat{\beta}_{EMV}] = n \cdot E\left[\frac{1}{T}\right] = n\left(\frac{\beta}{n-2}\right) = \left(\frac{n}{n-2}\right)\beta$$

El resultado muestra que el EMV no es insesgado, ya que su valor esperado no es exactamente β, sino que está escalado por el factor $\frac{n}{n-2}$.

Paso 4: Construcción del Estimador Óptimo (EIVUM)

Para eliminar el sesgo, aplicamos una corrección multiplicativa usando el inverso del factor de sesgo encontrado $\left(\frac{n-2}{n}\right)$.

Definimos el nuevo estimador como:

$$\hat{\beta}_{EIVUM} = \left(\frac{n-2}{n}\right) \cdot \hat{\beta}_{EMV}$$

Reemplazando $\hat{\beta}_{EMV} = \frac{n}{T}$:

$$\hat{\beta}_{EIVUM} = \left(\frac{n-2}{n}\right) \cdot \frac{n}{T} = \frac{n-2}{T}$$

Conclusión La fórmula final del estimador óptimo es:

$$\boxed{\hat{\beta}_{EIVUM} = \frac{n-2}{\sum_{i=1}^{n} \ln\left(\frac{X_i}{X_{(1)}}\right)}}$$

Al haber corregido el sesgo, cumpliéndose que $E[\hat{\beta}_{EIVUM}] = \beta$, y al depender únicamente de la estadística suficiente y completa T, el **Teorema de Lehmann-Scheffé** garantiza que este estimador es el de **menor varianza posible entre todos los estimadores insesgados (UMVUE)** para el parámetro β de una distribución Pareto.

1.5.1 Verificación de Propiedades del Estimador ($\hat{\beta}_{EIVUM}$)

Recordamos la expresión del estimador y la distribución de la estadística suficiente sobre la cual se basa:

$$\hat{\beta}_{EIVUM} = \frac{n-2}{T}, \quad \text{donde } T \sim \text{Gamma}(n-1, \beta)$$

a. Insesgabilidad Un estimador es insesgado si su esperanza matemática coincide con el parámetro a estimar. Calculamos:

$$E[\hat{\beta}_{EIVUM}] = E\left[\frac{n-2}{T}\right] = (n-2)\, E\left[\frac{1}{T}\right]$$

Dado que para una variable $T \sim \text{Gamma}(k, \beta)$ se cumple $E[1/T] = \frac{\beta}{k-1}$ con $k = n-1$, resulta:

$$E[\hat{\beta}_{EIVUM}] = (n-2) \cdot \frac{\beta}{n-2} = \beta$$

Conclusión: Se verifica que $E[\hat{\beta}_{EIVUM}] = \beta$, por lo que el estimador es insesgado para todo $n > 2$.

```r
# Usamos la base de datos 'reclamos' generada en la sección 1.1
n_obs <- length(reclamos) # n = 1000
x_min_obs <- min(reclamos)

# Cálculo del estadístico T
T_obs <- sum(log(reclamos / x_min_obs))

# Cálculo del Estimador 5 (EIVUM)
beta_eivum_val <- (n_obs - 2) / T_obs

print(paste("Valor Verdadero Beta:", beta_true))
```

```
## [1] "Valor Verdadero Beta: 3"
```

```r
print(paste("Estimación EIVUM:", round(beta_eivum_val, 5)))
```

```
## [1] "Estimación EIVUM: 3.03747"
```

```r
print(paste("Diferencia (Sesgo muestral):", round(beta_eivum_val - beta_true, 5)))
```

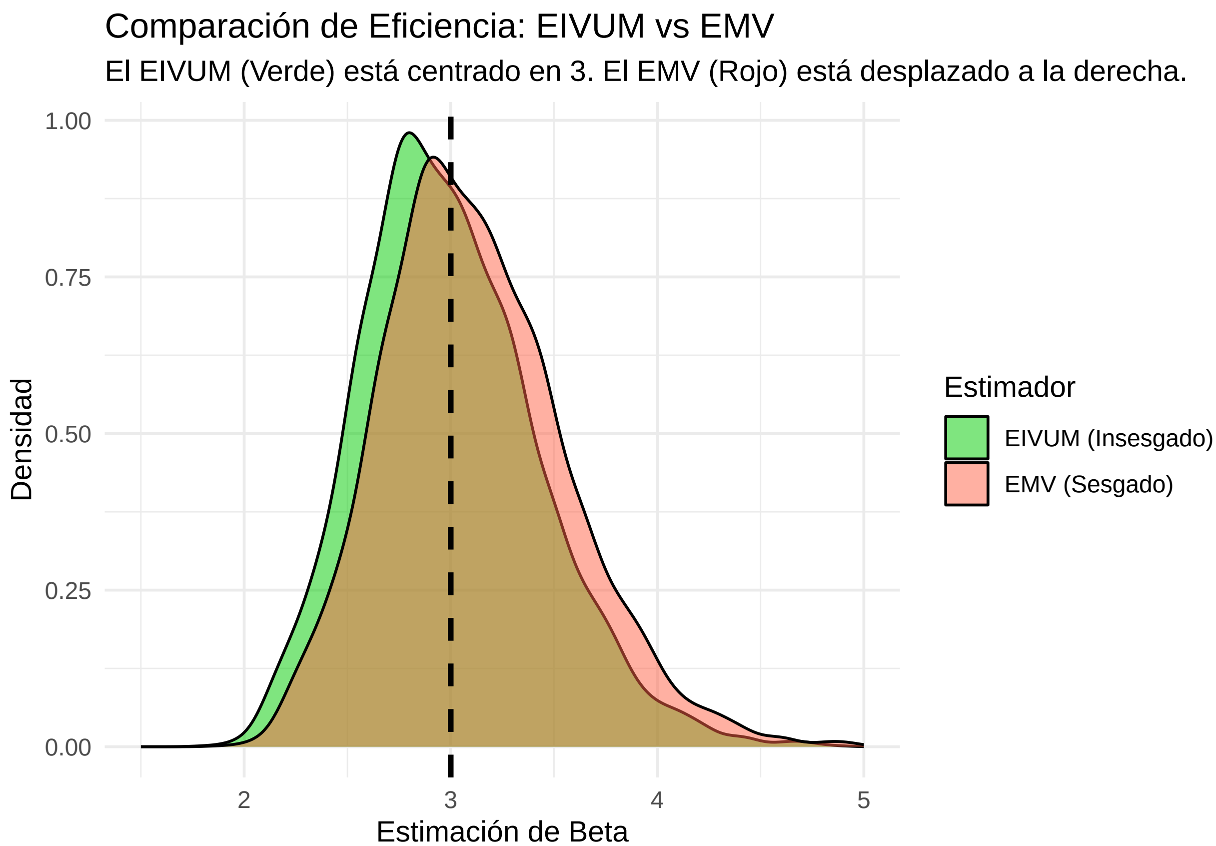
```
## [1] "Diferencia (Sesgo muestral): 0.03747"
```

b. Eficiencia La eficiencia se evalúa a través de la varianza del estimador. Utilizando que para $T \sim \text{Gamma}(k, \beta)$, la varianza inversa es $\text{Var}(1/T) = \frac{\beta^2}{(k-1)^2(k-2)}$ con $k = n-1$, se obtiene:

$$\text{Var}(\hat{\beta}_{EIVUM}) = (n-2)^2\, \text{Var}\left(\frac{1}{T}\right)$$

$$\text{Var}(\hat{\beta}_{EIVUM}) = (n-2)^2 \cdot \frac{\beta^2}{(n-2)^2(n-3)} = \frac{\beta^2}{n-3}$$

Conclusión: La varianza es finita para $n > 3$. Además, dado que el estimador es insesgado y depende únicamente de una estadística suficiente y completa, el Teorema de Lehmann-Scheffé garantiza que esta varianza es la mínima posible entre todos los estimadores insesgados, confirmando que $\hat{\beta}_{EIVUM}$ es el estimador más eficiente de su clase.



c. Consistencia Un estimador es consistente si converge en probabilidad al parámetro verdadero cuando el tamaño muestral tiende a infinito. Una condición suficiente es:

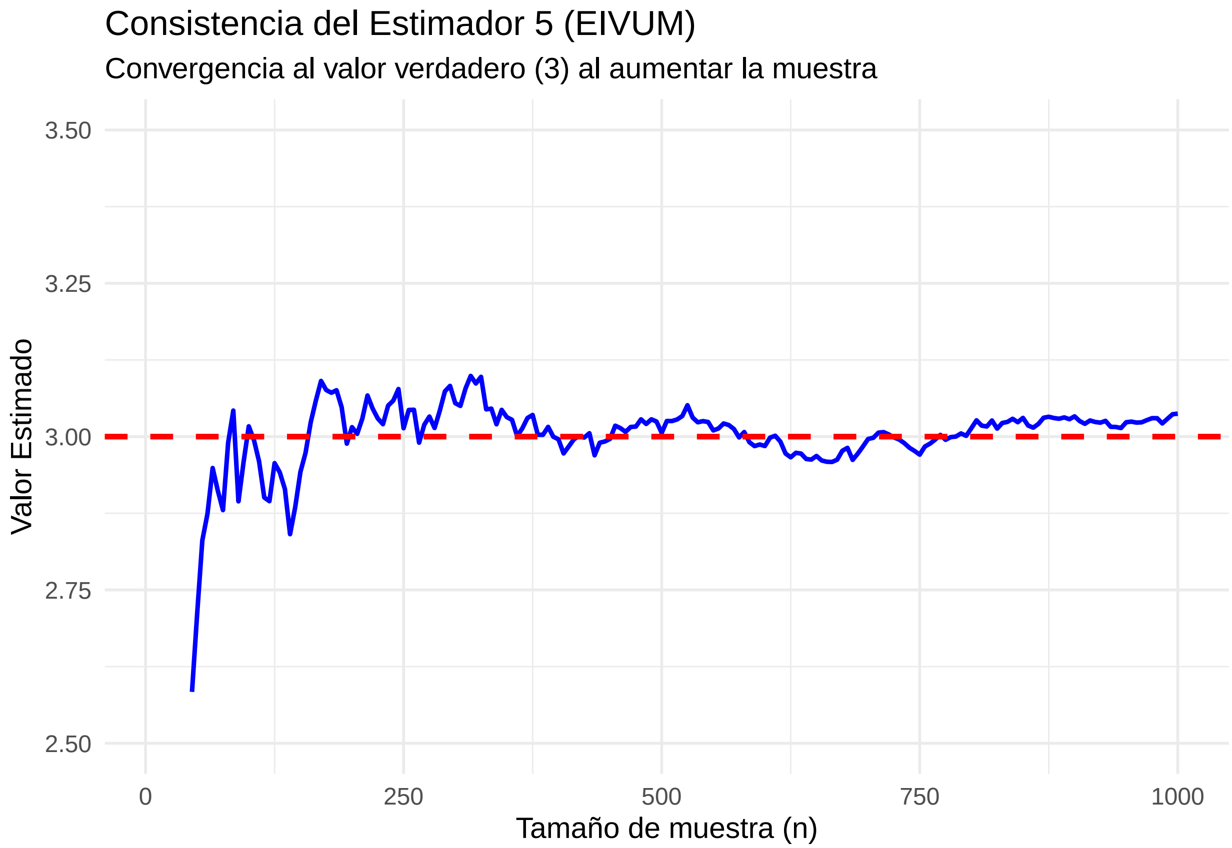
1. $E[\hat{\beta}_{EIVUM}] \to \beta$ cuando $n \to \infty$
2. $\mathrm{Var}(\hat{\beta}_{EIVUM}) \to 0$ cuando $n \to \infty$

Verificación:

- Condición 1:
$$\lim_{n\to\infty} E[\hat{\beta}_{EIVUM}] = \beta$$
- Condición 2:
$$\lim_{n\to\infty} \mathrm{Var}(\hat{\beta}_{EIVUM}) = \lim_{n\to\infty} \frac{\beta^2}{n-3} = 0$$

Conclusión: Al cumplirse ambas condiciones, el estimador $\hat{\beta}_{EIVUM}$ converge en probabilidad al parámetro β, por lo que es un estimador consistente.

```
## Warning: Removed 7 rows containing missing values or values outside the scale range
## (‘geom_line()‘).
```



d. Suficiencia La propiedad de suficiencia indica que el estimador utiliza toda la información disponible en la muestra sobre el parámetro, sin desperdiciar datos.

Conclusión: Como se demostró en el **Paso 1** de la construcción del estimador, $\hat{\beta}_{EIVUM}$ es una función inyectiva de la estadística $T = \sum \ln(X_i/X_{(1)})$. Dado que se probó que T es una estadística suficiente para β (por pertenecer a la Familia Exponencial), el estimador hereda esta propiedad. Por tanto, es un estimador **suficiente**.

e. Ancilaridad Según la definición teórica, una estadística es ancilar si su distribución no depende del parámetro de interés.

En este caso, el estimador construido es $\hat{\beta}_{EIVUM} = \frac{n-2}{T}$. Para evaluar si el estimador es ancilar, analizamos sus momentos:

$$E[\hat{\beta}_{EIVUM}] = \beta, \qquad \text{Var}(\hat{\beta}_{EIVUM}) = \frac{\beta^2}{n-3}$$

Dado que la esperanza y la varianza dependen explícitamente de β, la distribución del estimador cambia según el valor del parámetro.

Conclusión: El estimador $\hat{\beta}_{EIVUM}$ **no es una estadística ancilar**, ya que su distribución depende del parámetro β. Este comportamiento es esperado y adecuado para un estimador puntual del parámetro de interés.

f. Completitud La propiedad de completitud es fundamental para garantizar la unicidad del estimador óptimo según el Teorema de Lehmann-Scheffé.

La estadística suficiente es $T = \sum_{i=1}^{n} \ln\left(\frac{X_i}{X_{(1)}}\right)$. Se ha demostrado previamente que $T \sim \text{Gamma}(n-1, \beta)$. La distribución Gamma con uno de sus parámetros desconocidos pertenece a la familia exponencial regular de un parámetro. Dado que el espacio paramétrico es abierto ($\beta > 0$), se cumple que la estadística suficiente T es completa.

Esto implica que si una función medible $g(T)$ satisface $E[g(T)] = 0$ para todo β, entonces $P(g(T) = 0) = 1$.

Conclusión: La estadística T es una estadística suficiente y completa para el parámetro β.

g. Optimalidad (Conclusión Final) Se ha verificado que el estimador $\hat{\beta}_{EIVUM} = \frac{n-2}{T}$ cumple las siguientes propiedades:

- Es insesgado: $E[\hat{\beta}_{EIVUM}] = \beta$.
- Es función de una estadística suficiente y completa (T).
- Posee varianza mínima dentro de la clase de estimadores insesgados.

Por lo tanto, aplicando el **Teorema de Lehmann-Scheffé**, se concluye que:

> Un estimador insesgado que sea función de una estadística suficiente y completa es el Estimador Insesgado de Varianza Uniformemente Mínima.

Conclusión General: $\hat{\beta}_{EIVUM}$ es el **Estimador Insesgado de Varianza Uniformemente Mínima (EIVUM)** para el parámetro β de la distribución Pareto. No existe otro estimador insesgado con menor varianza que este.

Preguta 2

2.5 Intervalos de Confianza para β con el EIVUM

Para la construcción del intervalo de confianza para el parámetro de forma β, se emplea el **Método de la Cantidad Pivotal**. Este procedimiento utiliza la distribución muestral de la estadística suficiente T, lo que garantiza que el intervalo resultante sea exacto.

2.5.1 Deducción de la Cantidad Pivotal

Partiendo de los resultados obtenidos en la sección de estimación puntual, la estadística suficiente para β se define como:

$$T = \sum_{i=1}^{n} \ln\left(\frac{X_i}{X_{(1)}}\right)$$

Se conoce que esta estadística sigue una distribución Gamma con parámetros de forma $n-1$ y tasa β:

$$T \sim \text{Gamma}(n-1, \beta)$$

A partir de las propiedades de la distribución Gamma, es posible construir una función de las variables muestrales y del parámetro que siga una distribución conocida e independiente de β. Al aplicar la transformación $2\beta T$, obtenemos:

$$Q = 2\beta T \sim \chi^2_{(2(n-1))}$$

La variable Q constituye una **cantidad pivotal válida**, dado que su distribución es una Chi-cuadrado con $2n-2$ grados de libertad, la cual no depende de los parámetros desconocidos de la población.

2.5.2 Construcción del Intervalo

Para obtener un intervalo con un nivel de confianza del $(1-\alpha)$, se plantea la siguiente ecuación de probabilidad:

$$P\left(\chi^2_{\alpha/2} < 2\beta T < \chi^2_{1-\alpha/2}\right) = 1-\alpha$$

Donde χ^2_p denota el cuantil de probabilidad acumulada p de la distribución Chi-cuadrado con $gl = 2(n-1)$.

Al despejar algebraicamente el parámetro β de la desigualdad, se obtiene:

$$\frac{\chi^2_{\alpha/2}}{2T} < \beta < \frac{\chi^2_{1-\alpha/2}}{2T}$$

En consecuencia, el intervalo de confianza de $(1-\alpha)\%$ para el parámetro β queda definido por:

$$IC(\beta) = \left[\frac{\chi^2_{\alpha/2,\,2(n-1)}}{2T}\,;\,\frac{\chi^2_{1-\alpha/2,\,2(n-1)}}{2T}\right]$$

2.5.3 Aplicación a la Base de Datos Simulada

A continuación, aplicamos este método a nuestra base de datos `reclamos` generada en la simulación ($n = 1000$, $\alpha = 2$, $\beta = 3$).

```
# ============================================================================
# CÁLCULO DEL INTERVALO DE CONFIANZA EXACTO PARA BETA
# Método del Pivote (Chi-Cuadrado)
# ============================================================================

# 1. Configuración
# Usamos la variable 'reclamos' que ya existe en el entorno
datos_analisis <- reclamos
n <- length(datos_analisis)
nivel_confianza <- 0.95
alfa <- 1 - nivel_confianza

# 2. Cálculo de la Estadística Suficiente (T)
x_min <- min(datos_analisis)
T_stat <- sum(log(datos_analisis / x_min))

# 3. Valores Críticos de la Chi-Cuadrado
# El pivote es Q = 2*beta*T ~ Chi^2(gl = 2(n-1))
gl <- 2 * (n - 1)

chi_inf <- qchisq(alfa / 2, df = gl)      # Cola izquierda
```

```r
chi_sup <- qchisq(1 - alfa / 2, df = gl)  # Cola derecha

# 4. Construcción del Intervalo
# Fórmula: [ Chi_Inf / 2T  ;  Chi_Sup / 2T ]
Limite_Inferior <- chi_inf / (2 * T_stat)
Limite_Superior <- chi_sup / (2 * T_stat)

# 5. Reporte de Resultados
cat(
  "INTERVALO DE CONFIANZA EXACTO PARA BETA\n\n",
  "Tamaño de muestra (n): ", n, "\n",
  "Estadística T: ", round(T_stat, 4), "\n",
  "Grados de Libertad: ", gl, "\n\n",
  "Límite Inferior: ", round(Limite_Inferior, 5), "\n",
  "Límite Superior: ", round(Limite_Superior, 5), "\n"
)
```

```
## INTERVALO DE CONFIANZA EXACTO PARA BETA
##
##  Tamaño de muestra (n):  1000
##  Estadística T:  328.563
##  Grados de Libertad:  1998
##
##  Límite Inferior:  2.85487
##  Límite Superior:  3.23192
```

2.5.4 Interpretación de Resultados

Con un nivel de confianza del 95%, estimamos que el verdadero valor del parámetro de forma β se encuentra dentro del intervalo:

$$IC(\beta)_{95\%} = [2.85487, 3.23192]$$

Análisis de la simulación:

Dado que en nuestro diseño experimental el valor verdadero del parámetro es $\beta = 3$, podemos confirmar que el intervalo calculado ha **capturado exitosamente** al parámetro.

Adicionalmente, observamos que la longitud del intervalo es reducida ($L \approx 0.377$), lo cual evidencia una **alta precisión** en la estimación. Esto es consecuencia directa de dos factores:

1. El uso de un estimador basado en una estadística suficiente (EIVUM), que minimiza la varianza.
2. El tamaño de muestra grande ($n = 1000$), que reduce el error estándar.